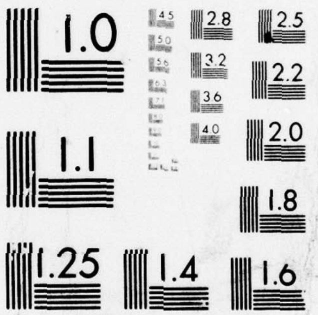
AD-A076 563    CARNEGIE-MELLON UNIV PITTSBURGH PA DEPT OF COMPUTER --ETC    F/G 6/4
WORKSHOP ON CONTROL STRUCTURES AND KNOWLEDGE REPRESENTATION FOR--ETC(U)
AUG 79    R REDDY , A ROSENFELD , J WESZKA    N00014-79-M-0009

UNCLASSIFIED    NL

1 OF 1
AD
A076563

END
DATE
FILMED
12-79
DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

# Final Report on Workshop on Control Structures and Knowledge Representation for Image and Speech Understanding

April 3-4, 1979
Center for Adult Education
University of Maryland
College Park, MD 20742

Sponsored by
National Science Foundation
Washington, D.C. 20550
Grant MCS-78-21471
and
Office of Naval Research
Arlington, VA 22217
Grant N00014-79-M-0009
to
Department of Computer Science
Carnegie-Mellon University
Pittsburg, PA 15213

Principal Investigators:
Raj Reddy
Carnegie-Mellon University
Azriel Rosenfeld
University of Maryland

Report prepared by
Joan Weszka
IBM Corporation

August 31, 1979

9 Final rept,

# Final Report on
6 Workshop on
# Control Structures and Knowledge Representation for Image and Speech Understanding

held on

April 3-4, 1979, at
Center for Adult Education, College Park Maryland.
University of Maryland
College Park, MD 20742

Sponsored by
National Science Foundation
Washington, D.C. 20550
Grant MCS-78-21471
and
Office of Naval Research
Arlington, VA 22217
15 Grant N00014-79-M-0009, NSF-MCS78-21471
to
Department of Computer Science
Carnegie-Mellon University
Pittsburg, PA 15213

11 31 Aug 79

Principal Investigators:
10 Raj Reddy,
Carnegie-Mellon University
Azriel Rosenfeld
University of Maryland

12 53

Report prepared by
Joan Weszka
IBM Corporation

August 31, 1979

403 081

1B

## Table of Contents

# 1. Introduction

Extensive research has been devoted to the development of systems for image and speech understanding. Both of these problem domains involve mappings from a complex input signal to a symbolic description, and make use of knowledge on many levels. As a result, workers in both domains have found it necessary to develop sophisticated control structures. Many of the ideas about knowledge representation and processing control that have evolved in one of these fields are applicable to the other. Unfortunately, there has been very little communication between the two groups of researchers.

The purpose of this workshop was to review the major ideas on control structures and knowledge representations that have developed in the image and speech understanding areas, and to provide an opportunity for cross-fertilization between the two fields. The presentations made at the workshop were intended to stress general aspects, rather than problem-domain-dependent considerations.

This report contains the workshop's program, a summary of each presentation, and a critical overall evaluation. A list of attendees is included in an Appendix. A bibliography has not been provided, but many papers on the representation and control aspects of image and speech understanding can be found in

A. R. Hanson and E. M. Riseman, Eds., Computer Vision Systems, Academic Press, New York, 1978.

W. A. Lea, Ed., Trends in Speech Recognition, Prentice-Hall, Englewood Cliffs, NJ, 1979.

## 2. Workshop Program

Introduction: A. Rosenfeld

Perspectives Session*

Speakers: D. R. Reddy
A. Rosenfeld

Issues Sessions*

Speakers: V. Lesser
J. Feldman
F. Hayes-Roth
H. Barrow
L. Erman
E. M. Riseman
J. K. Baker
R. Davis
R. M. Haralick

Concluding Remarks: P. H. Winston
A. Rosenfeld

* The presentations in these sessions are summarized in this report.

## 3. Perspectives Session

### a. D. R. Reddy

Reddy's opening remarks set the tone for the workshop. Its purpose was to promote cross-fertilization in the speech and vision areas--to communicate valuable insights in one area that might be useful in solving problems in the other area.

AI is the "study of how to use knowledge to achieve intelligent action, which often implies selection from a large space of alternatives." Vision and speech are two problems which require the application of diverse sources of knowledge, including both symbolic knowledge and knowledge of the signal space, to the interpretation of a noisy signal (image or speech waveform). AI systems which solve vision and speech problems differ from purely symbolic problem solving systems since they must explicitly deal with errors, noise, and uncertainty in the input data. Furthermore, systems for analyzing speech and images have a natural hierarchical structure. Figure 1 shows the hierarchical structure inherent in speech understanding.

The role of knowledge in solving problems is to constrain the alternatives at each stage of the search process in order to reduce exponential growth. The relationship between knowledge and search is that "knowledge reduces uncertainty and, therefore, search, and conversely, search can compensate for a lack of knowledge."

A key issue in AI is representation. From an informal, common sense point of view, a represenation may be viewed as "a set of conventions about how to describe things." Examples of representation systems include:

1) algebra: $(X + Y)(X - Y) = X^2 - Y^2$

2) logic: $P\wedge(Q\vee R) = (P\wedge Q)\vee(P\wedge R)$

3) grammar: NP + VP → S

4) productions: (Goal=x)(State=y) → (Subgoal M'th=Y)

5) pattern directed languages: PLANNER, CONNIVER, QA4

More formally, a representation may be defined as "a set of data structures that make a body of knowledge available to a processor," or, more graphically:

Representation =

Content + Accessibility +

Augmentability + <Other properties of memory>

Turning to the problem of control, the following three questions address central issues involved in the selection of control structures:

1) How do different knowledge sources interact in solving a problem?

2) How does one generate alternative paths (solutions) in a large combinatorial search space?

3) How does one focus attention on some subset of the search space?

The problem of how knowledge sources interact can be answered by describing knowledge sources as filters. Each knowledge source acts as a filter, thus constraining the set of alternatives which must be considered by other knowledge sources. One can apply different knowledge sources either successively, each operating on the previous filtered output, or in parallel. In order for a knowledge source to be applicable, however, the alternatives must be specified in a domain that is consistent with the knowledge source. Models for knowledge source interaction include:

1) hierarchical models (bottom-up filtering),

2) goal-directed models (top-down filtering),

3) heterarchical models (each knowledge source communicates with all others),

4) blackboard models (knowledge sources "post" hypotheses which may be accepted, modified or rejected by other knowledge sources),

5) integration model (integration of knowledge sources into a unified homogeneous representation).

The question of how to generate alternative solutions in a search space may be answered by a number of different search techniques. The most appropriate technique for a given problem depends on the requirements of the solution (e.g., is an optimal solution sought?). The following alternatives are available:

1) A satisfying search can be achieved using depth-first, breadth-first, or hill-climbing strategies.

2) An optimal search can be achieved using best-first, dynamic programming, branch and bound, A*, or short-fall density. If a near-optimal solution is satisfactory, then search cost and effort can be saved by using beam search.

3) An "information gathering search" (constraint satisfaction) can be achieved using Waltz filtering, production systems, relaxation, or the Hearsay model.

The question of how to focus attention on some subset of the search space is basically a question of deciding what to do next. Decisions can be based on goodness estimates of hypotheses (hill-climbing, best-first with backtracking, or best-few as in beam search) or on divide and conquer strategies (anchor points and islands of reliability).

The designers of control structures and knolwedge representations for new image and speech understanding systems must make many other crucial design decisions. Some of these decisions include:

1) conventional search vs. information gathering search,

2) compiling vs. interpreting knowledge,

3) balancing the cost of decision making vs. the cost of executing the decision,

4) explicit representation vs. implicit representation of hypotheses,

5) integrated vs. independent knowledge sources,

6) fact distribution over a large number of small bodies of knowledge vs. fact distribution over a small number of large bodies of knowledge (i.e., granularity of knowledge source decomposition).

Abstraction plays an important role in speech and visual understanding systems. Abstraction can be categorized into three areas:

1) Signal abstraction, exemplified by Kelly's planning and pyramid representations of images. In the continuous domains, one can find space, time and area abstractions.

2) Knowledge source abstraction, exemplified by the abstraction of object descriptions (Binford), the abstraction of maps (Thorndyke), and the abstraction of rules (Hayes-Roth, Waterman, Klahr and Rubin).

3) Control strategy abstraction, exemplified by scene, viewpoint and structure identification (Rubin), hierarchical beam search, and hierarchical relaxation.

There are many other issues. One of the most important is that image or speech understanding can be viewed as a large number of loosely coupled subproblems. This means that many of the relevant contextual constraints are utilized within the first few steps of processing. Good examples of this behavior can be seen in beam search and relaxation processes.

A second issue is problem space decomposition, especially the discovery of anchor points (points of locally high certainty) in both the signal space and the knowledge space. The discovery of such points allows the search problem to be decomposed into relatively independent subproblems, which makes the search for the solution easier. However, finding the anchor points is iself a difficult search problem, a fact which is often overlooked.

The following areas appear to be the most important areas for future research:

1) knowledge acquisition,

2) dynamic adaptation of systems,

3) problems in dealing with large bodies of knowledge, and

4) graceful interaction of systems with regard to the addition of new objects and new structures.

b. A. Rosenfeld

Rosenfeld opened his talk by discussing a prevalent image analysis paradigm: An image is first segmented into parts which are then described by their shapes, textures, and other properties. These parts are then integrated into a relational network model of the structure of the image, which is then matched against known, previously stored models during a recognition process.

Segmentation strategies are ordinarily either parallel or sequential. Parallel strategies are order-independent and can be implemented on fast processors, but are limited in intelligence by their reliance on local decisions. Sequential strategies, on the other hand, are order-dependent and slow, but are more intelligent since they operate on a "learn-as-you-go" basis.

Relaxation techniques, which are iterative, represent an alternative strategy. Their advantages are that they are nearly as fast as parallel strategies (since each iteration is done in parallel), they are increasingly smart (since successive iterations make use of decisions at previous iterations), and they defer commitment by making fuzzy or probabilistic decisions at each stage. At the University of Maryland, an image analysis system has been constructed which reads handwritten words using relaxation methods. This system is, in fact, a hierarchical relaxation system. It would be desirable to develop a similar approach to speech understanding and to compare the results with those obtained using existing speech understanding systems.

There are still many open problems in image analysis that need to be studied. In the area of segmentation, research is needed on subjects including statistical models for describing texture and geometric models for describing shapes. Problems in representation include the pervasive problem of performing the transformation from the image data structure to a graph data structure, and the fundamental process of devising methods for bringing high-level knowledge to bear on that transformation. Pyramid data structures may be very useful in solving these problems.

The following approach, based on relaxation, represents a powerful solution to the recognition (i.e., model matching) problem:

1) Consider all pairings of model nodes with actual nodes representing image parts.

2) Assign initial confidences based on properties of the image nodes.

3) Adjust these confidences, using a relaxation system, based on relations between the nodes and the confidences of the related nodes.

Experimentally, the initial ambiguity in matching model nodes with actual nodes is greatly reduced after only a few iterations of the relaxation process.

## 4. Issues Session

### a. V. Lesser

Lesser's talk focused on issues of control in knowledge-based systems that handle uncertainty in both data and control. Examples of such systems are MYCIN, PROSPECTOR, relaxation systems, MSYS, HEARSAY II, HARPY and HWIM. The tasks undertaken by these systems have the following characteristics:

1) complexity of the search space,

2) granularity of knowledge (cost of knowledge application),

3) uncertainty in input data and knowledge (degree and type),

4) distribution of information in data (uniform vs. clustered),

5) coupling (interdependency) of partial solutions.

The control structures for existing knowledge-based systems can be described according to the following design alternatives:

1) opportunistic (data-directed) vs. fixed control,

2) decentralized vs. centralized,

3) optimal vs. heuristic,

4) multi-level vs. single-level.

The comparison of existing systems is, unfortunately, very difficult. No models for the comparison of systems from the point of view of search strategies have been developed. The problem of how to evaluate system performance must also be addressed. Comparing two complex systems is very difficult. One can either try to adopt formal (e.g., Bayesian) methods for comparison, or simply rely on ad hoc, heuristic methods.

The following open issues in knowledge-based system design must be addressed:

1) How should alternatives be represented? Choices include integrated or separate contexts.

2) How should knowledge be structured? Choices include appropriate and cheap structures vs. precise and expensive ones, and anonymous and independent vs. integrated knowledge.

3) How should focusing knowledge be integrated? Choices include task independent vs. task dependent integration, and local vs. global integration.

These issues suggest the following topics for study:

1) Using resource constraints to select appropriate search strategies;

2) developing models for the sensitivity of systems to uncertainty, new data, etc.;

3) decomposing systems into components whose control structures are well understood and identifiable;

4) developing methodologies for building systems which can be compared;

5) designing test environments for evaluating system performance; and

6) comparing speech and vision systems--e.g., relaxation systems vs. the HEARSAY system.

b. J. Feldman

Feldman presented a methodology for the creation of a query-directed image understanding system based on the point of view that perception is "the maximization of the expected utility to the perceiver." This system has been used to solve a variety of tasks, including the detection of ribs in chest x-rays. There are three levels of data structure in the system (see Figure 2):

1) an image data structure containing information gathered in a non-purposive way,

2) a sketch map containing knowledge used in a particular application, and

e) a model containing information in a knowledge data base.

The knowledge data base contains information concerning where and how to look, how to model objects, and how to fit the models to data. The knowledge is stored as a set of procedures, each of which has associated with it a procedure descriptor which describes the procedure declaratively. Thus, one procedure may "reason" about another procedure's capabilities and performance in the model. The contents of a procedure descriptor includes its cost, confidence or reliability, resource requirements, pre-conditions and post-conditions.

c. F. Hayes-Roth

Hayes-Roth addressed the problem of hierarchies in interpretation systems. He gave examples of such hierarchies, and discussed the functions they support as well as the roles they play in existing systems.

The functions of a hierarchy as exemplified by the HEARSAY II system include simplification of the knowledge base, allowing for the sharing of intermediate results and aggregation of partial results, exploitation of specialized knowledge, and allocation of resources for processing promising data. Hierarchies play various roles in the areas of knowledge acquisition, inference, hypothesization, communication and control. These functions are reflected in the structure of problems as diverse as solving message puzzles using many communicating processes, each with a partial view of a dynamic puzzle, and simulating and monitoring a complex activity such as tactical and strategic troop and vehicle movements.

The roles of hierarchies, then, include:

1) knowledge acquisition--abstraction, induction, generalization, specialization, intelligibility, modifiability.

2) inference--simplification, aggregation, approximation, intelligent search, constraint satisfaction.

3) hypothesization--data abstraction, exploitation of partial results, modeling situations, prediction, surprise detection.

4) communication--linguistic abstraction, context-dependent coding, expectation-filtering, network simplification.

5) control--process and performance abstractions, complexity modulation, focus, resource allocation, (global) coordination.

Hierarchies in a knowledge base system can be viewed in the following ways:

1) knowledge is a conceptual hierarchy,

2) inference is a power/performance hierarchy,

3) hypotheses reflect a quality/consistency hierarchy,

4) communication is a contextual hierarchy, and

5) control is a pragmatic hierarchy.

d. H. Barrow

Barrow's talk focused on two themes: the recovery of intrinsic information from grey level images, and the utilization of high-level knowledge to disambiguate the interpretation of regions in an image.

Vision systems can be divided into two major parts:

1) a low level where processing is parallel and data-driven, and the representation of information is iconic, or picture-like;

2) a high level where processing is serial and goal-directed, and the representation of information is symbolic.

Figure 3 shows this division, along with the different types of information and knowledge available at each level.

A great deal of low-level information about objects can be recovered from images if one can make certain assumptions about the nature of the objects in the scene and the illumination of the scene. In particular, consider a simple world where objects have smooth surfaces, uniform reflectance, and no planar surfaces or sharp corners, and where illumination is provided by a distant point source and a uniform background (e.g., sky) illumination. A vision system can derive "intrinsic" characteristics of image points (such as distance to the object point corresponding to any image point) based on the object and illumination models. In general, the edges in the image give rise to important clues as to the values of intrinsic properties at the boundaries, which might then be propagated into the interiors of objects. Much more work needs to be done before intrinsic properties can actually be computed for even simple scenes.

The central question at the high level is how knowledge can be brought to bear on the disambiguation of scene parts. A system, called MSYS, was developed at SRI for this purpose. In MSYS, information about scenes is of two types:

1) Information about object properties, such as size, color, etc., that may be found in scenes. This information is used to assign initial likelihoods to interpretations of scene parts.

2) Information about the spatial relations between objects in the scene, expressed in the form of real valued constraint relations (e.g., a door may be found above a floor with a likelihood of .7). MSYS finds a best overall interpretation of the regions in the scene by integrating a "relaxation-like" constraint satisfaction procedure into a standard ordered search procedure.

As a supplement to Barrow's talk, J. M. Tenenbaum presented a detailed example of the application of MSYS to a scene interpretation problem.

e. <u>L. Erman</u>

Erman's talk focused on the design of HEARSAY III. He discussed various concepts underlying its design. These included "aggregates" which are sequence hypotheses or "AND" nodes, and "hypotheses" which correspond to "OR" nodes. General context mechanisms and a flexible scheduling algorithm are also important in HEARSAY III. Figures 4-8 list some relevant design issues concerning aggregates, hypotheses, contexts, scheduling, and knowledge source activation. It should be pointed out that in building a system such as HEARSAY III, there are trade-offs in generality, efficiency and naturalness.

HEARSAY III has developed from early work on production systems. Figure 9 is a block diagram of the components of a Production System, and Figure 10 shows an analogous diagram for HEARSAY III. The key differences are:

1) the substitution of general knowledge sources for the simple set of production rules with actions,

2) the upgrading of the workspace to a structured data base and then to a blackboard,

3) the substitution of a powerful scheduling algorithm for the simple conflict resolution scheme, and

4) the replacement of the simple pattern matcher with a general knowledge source evaluation procedure.

In summary, the design goals of HEARSAY III are:

1) to identify and supply basic domain independent mechanisms and to leave policy decisions to the user,

2) to improve on methods of representation on the blackboard, and

3) to develop the system to the point where its competence and performance can be evaluated separately.

f. E. M. Riseman

Riseman discussed the organization of the vision system constructed at the University of Massachusetts. There are three levels of representation in this system--long-term memory, short-term memory, and a memory for regions, segments and vertices which describe the results of scene segmentation. The task of image interpretation is divided into two major subtasks:

1) instantiation of a relevant schema drawn from long-term memory;

2) top-down interpretation of the image according to that schema. This interpretation utilizes the various knowledge sources of the system--curve fitting, shape analysis, spectral analysis, etc.

Figure 11 shows the structure of parts of long-term and short-term memory during the analysis of a simple outdoor scene.

g. J. K. Baker

Baker discussed the design of speech understanding systems. Very powerful systems can be built based on Markov models of the speech generation process, and a dynamic programming analysis of the speech waveform based on this Markov model. The advantages of this approach are that the resulting speech systems are both easy to implement and easy to train. One such system developed at IBM achieved accuracy levels of 95% on individual word recognition, using a 250 word task and high quality speech input.

The design of these speech systems should also be of interest to researchers in vision, because the dynamic programming used to analyze speech bears some resemblance to the relaxation techniques prevalent in image understanding.

h. R. Davis

Davis spoke on the problems involved in the interactive transfer of expertise in knowledge-based systems. Such systems are characterized by a large amount of knowledge, much of which is ill-defined.

Problems with large knowledge bases arise because not all of the knowledge can be specified in one step. Often, many iterations are required to enter the initial knowledge base and, over time, many updates are required. These problems could be alleviated by developing an environment which facilitates making changes to the knowledge base, and by improving its comprehensibility. This can be done by making the knowledge base modular so as to limit propagation effects and, at the same time, "self-adjusting" so that propagation effects can be automated.

The problem of ill-defined knowledge arises because new knowledge entered into the system will probably be wrong the first time, often because it is too general. It would be desirable to build "forgiving" systems in which some level of performance is possible, even with incomplete or incorrect knowledge, and in which approximate knowledge can be used. These are areas in which future research is needed.

i. R. M. Haralick

Haralick's presentation addressed the computational savings associated with using certain look-ahead, or discrete relaxation, operators in search. He compared his results with those of Gaschnig, who found that for generalizations of the 8-queens problem, some forms of backtrack programming were more efficient than the combination of traditional backtrack programming with discrete relaxation.

Different results are obtained when one considers another set of problems which are abstract labeling programs--i.e., one is given a set of objects and labels, along with a constraint relation which specifies which pairs of labels may be associated with which pairs of objects. The goal is to assign labels to objects such that all pairs of object-label associations simultaneously satisfy the constraints. More formally, one is given

1) a set U of units,

2) a set L of labels,

3) a cover C for U, and

4) a constraint set, $R(C) = \{f | f:C \rightarrow L\}$

The goal is to find all mappings h from U to L such that for every $c \in C$ there exists $f \in R(C)$ satisfying $U \in C \Rightarrow f(u)=h(v)$.

For this class of problems the combination of backtrack programming and discrete relaxation was more efficient than sophisticated backtrack programming alone. Some mathematical analysis also supports these conclusions.

## 5. Concluding Remarks

### a. P. H. Winston

Winston expressed his disappointment that so many researchers seemed to be stressing problems concerning control at the expense of studying the harder, but more important, problems associated with representation. The research on vision at M.I.T. has focused on developing representations for visual tasks, such as stereo vision, texture and motion analysis. Once adequate representations are developed, the choice of control structure becomes a secondary, and usually trivial, problem. As a concrete example, Waltz's thesis was essentially concerned with the representation problem for very general blocks-world scenes containing shadows, cracks, etc. Once Waltz developed this representation, then the choice of a control structure for actually interpreting a scene was straightforward.

b. A. Rosenfeld

Rosenfeld closed the workshop by speculating that there will be no general theory of representation or control in vision and speech until some basic AI prejudices are overcome. These prejudices include:

1) generality, which has led to a preoccupation with philosophy, and a proliferation of heuristic ideas which don't map well into mathematical models;

2) symbolic computation, which has led to a reluctance to use fuzzy or probabilistic methods (an important class of exceptions are the expert systems such as MYCIN or PROSPECTOR);

3) goal-directedness, which has led us to think too hard about processes that in humans are largely non-conscious; and

4) search, which keeps us trying to make sequential processes smarter and faster. It is not clear that even relatively high levels of analysis can't be parallel.

It was pointed out, however, that workers in image and speech understanding are probably less subject to these prejudices than any other group of AI researchers, so that the danger of adhering to them unnecessarily should be minimal.

## 6. Summary

### a. Why this workshop?

Speech and image understanding have many common aspects. Both involve the description, usually in natural language, of a complex input signal, which may be noisy and distorted. Both seem to lend themselves to hierarchical treatment; examples of such hierarchies were given by Reddy, Barrow, and Riseman, among others. They do differ in that speech is produced with intent to communicate, whereas vision deals with arbitrary natural scenes; but general "sound understanding" is analogous to general vision, and certain specific vision domains (e.g. handwriting) are analogous to speech.

The high degree of commonality suggests that useful cross-fertilization should be possible between the speech and image understanding communities. Some has indeed taken place, notably in Reddy's work; but communication between the two groups is still rather limited--they publish in different journals, attend different meetings, etc. Hopefully this workshop has served to strengthen the ties between the two fields.

A more important purpose of the workshop was to focus on the general issues that are common to both image and speech understanding--issues of knowledge representation and control. By comparing disparate viewpoints on these common issues, we should be able to achieve a greater degree of insight and understanding (a "stereoscopic perspective", to stretch the metaphor), leading ultimately to the development of a theory of representation and control in these and similar problem domains.

b. What do we mean by a theory? Why don't we have one?

On a general level, an image or speech understanding system is defined by specifying

a) The ensemble of inputs to be analyzed

b) The ensemble of possible descriptions

c) The relevant knowledge base

d) The data structures to be used to represent the input data, the given knowledge, and the intermediate derived information

e) The strategies (search, parallelism, etc.) used by the procedures which operate on these data structures to generate the desired description.

Ideally, a theory of speech or image understanding should provide us with the ability to quantitatively predict (or at least give bounds on) the expected performance of a system, once the system has been specified.

Experience with a variety of systems over the past decade has given us many insights about the performance of various approaches. It has become easy for us to formulate qualitative taxonomies of representations and control structures on various levels; several good ones were in fact presented at the Workshop (by Reddy, Lesser, etc.). Many useful general-purpose strategies have also been formulated (e.g., MSYS/relaxation, beam search, etc.). Some systems have already passed through several generations of evolution; an example is the progression from production systems to HEARSAY III, as reviewed by Erman.

What is still lacking, however, is a quantitative theory. Unfortunately, for most nontrivial problem domains, the tasks of fully modelling the input data, the knowledge base, and the expected control structure performance are mathematically intractable. Moreover, researchers are primarily concerned with designing and building working systems, and have not devoted the effort that would be required to develop a mathematical theory of such systems. In fact, most AI researchers are not inclined toward working on mathematical theories, since such theories are likely to be of limited scope at best.

### c. What can we do now?

If we really want to move in the direction of a theory of speech and image understanding, we can begin by defining *minidomains* for which the necessary modeling tasks will be tractable. This involves specifying simple input data ensembles and knowledge bases (or perhaps simplified models for complex ones), and choosing simple data structures and control strategies (which, for the sake of tractability, may have to be nonhierarchical, nonparallel, and generally trivial by contemporary standards). Such minidomains may be unimpressive in comparison with the domains handled by today's working systems, but they will provide us with vital experience in quantitative modeling, and will eventually lead toward predictive theories capable of handling real-world situations.

At the same time, we can begin to build up a *descriptive* science of image and speech understanding by carrying out *quantitative performance evaluations* of existing and future systems. Wherever possible, common input data should be used, so that comparative evaluation is possible. Exchanges of input data and performance comparisons will not only provide a base of experimental data for future theoretical analysis, but will also lead to increased cooperation and collaboration among researchers in the field.

## APPENDIX: ATTENDANCE LIST

A. Invited Participants:

James K. Baker
Dialog Systems
Belmont, MA 02172

Harry G. Barrow
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Larry S. Davis
Department of Computer Science
University of Texas
Austin, TX 78712

Randall Davis
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square
Cambridge, MA 02139

Lee Erman
ISI
University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90291

Jerome S. Feldman
Department of Computer Science
University of Rochester
Rochester, NY 14627

K. S. Fu
School of Electrical Engineering
Purdue University
West Lafayette, IN 47907

Robert M. Haralick
Department of Electrical Engineering
Virginia Polytechnic Institute & State University
Blacksburg, VA 24061

Frederick Hayes-Roth
Rand Corporation
1700 Main Street
Santa Monica, CA 90406

Victor Lesser
Department of Computer & Information Science
University of Massachusetts
Amherst, MA 01003

David L. Milgram
Lockheed Palo Alto Laboratories
D52-53, Building 204
3251 Hanover Street
Palo Alto, CA 94304

D. R. Reddy
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213

Edward M. Riseman
Department of Computer Science
University of Massachusetts
Amherst, MA 01002

Azriel Rosenfeld
Computer Science Center
University of Maryland
College Park, MD 20742

Steven Rubin
Bell Laboratories
Crawfords Corner Road
Holmdel, NJ 07733

J. M. Tenenbaum
SRI International
333 Ravenswood Avenue
Menlo Park, CA 94025

Joan S. Weszka
IBM Corporation
11400 Burnet Road
Austin, TX 78759

Patrick H. Winston
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square
Cambridge, MA 02139

Steven W. Zucker
Department of Electrical Engineering
McGill University
Montreal, Canada H3A 2A7

B. Government Observers:

Eamon B. Barrett
National Science Foundation
1800 G Street, N.W.
Washington, D.C. 20550

Henry R. Cook
Defense Mapping Agency
6500 Brookes Lane
Washington, D.C. 20315

Judith Davenport
Defense Mapping Agency
6500 Brookes Lane
Washington, D.C. 20315

John K. Dixon
Computer Science Laboratory
Naval Research Laboratory
Washington, D.C. 20375

Larry Druffel
DARPA/IPTO
1400 Wilson Boulevard
Arlington, VA 22209

Robert S. Engelmore
DARPA/IPTO
1400 Wilson Boulevard
Arlington, VA 22209

Gordon D. Goldstein
Office of Naval Research
Information Systems Branch
Arlington, VA 22217

R. A. Kirsch
National Bureau of Standards
Washington, D.C. 20234

Don J. Orser
Center for Applied Mathematics
National Bureau of Standards
Washington, D.C. 20234

C. Other Observers:

Raj Aggarwal
Honeywell, Inc.
2600 Ridgway Parkway
Minneapolis, MN 55413

Jonathan Allen
Artificial Intelligence Laboratory
Massachusetts Institute of Technology
545 Technology Square
Cambridge, MA 02139

A. J. Cote, Jr.
Applied Physics Laboratory
Johns Hopkins University
Johns Hopkins Road
Laurel, MD 20810

Jan-Olof Eklundh
Computer Science Center
University of Maryland
College Park, MD 20742

Robert Futrelle
Department of Genetics & Development
University of Illinois
Urbana, IL 61801

Ramesh Jain
Department of Computer Science
Wayne State University
Detroit, MI 48202

Hans-Hellmut Nagel
Fachbereich Informatik
Universität Hamburg
Schlüterstrasse 70
2000 Hamburg 13/Germany

Bernd Neumann
Fachbereich Informatik
Universität Hamburg
Schlüterstrasse 70
2000 Hamburg 13/Germany

Bernd Radig
Fachbereich Informatik
Universität Hamburg
Schlüterstrasse 70
2000 Hamburg 13/Germany

Linda Shapiro
Department of Computer Science
Virginia Polytechnic Institute
Blacksburg, VA 24051

Michael Shneier
Computer Science Center
University of Maryland
College Park, MD 20742

Flavio R. D. Velasco
Computer Science Center
University of Maryland
College Park, MD 20742

Angela Wu
Computer Science Center
University of Maryland
College Park, MD 20742

D. Students:

Narendra Ahuja
Computer Science Center
University of Maryland
College Park, MD 20742

S. T. Barnard
Department of Computer Science
University of Minnesota
Minneapolis, MN 55455

Bernd Bruegge
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213

Alan Danker
Computer Science Center
University of Maryland
College Park, MD 20742

Charles Dyer
Computer Science Center
University of Maryland
College Park, MD 20742

Mark Fox
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213

Kenneth Hayes
Computer Science Center
University of Maryland
College Park, MD 20742

Martin Herman
Computer Science Center
University of Maryland
College Park, MD 20742

Tsai-Hong Hong
Computer Science Center
University of Maryland
College Park, MD 20742

David Hornig
353 Stratford Avenue
Pittsburgh, PA 15232

Robert Hummel
Department of Mathematics
University of Minnesota
Minneapolis, MN 55455

John R. Kender
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213

Robert L. Kirby
Computer Science Center
University of Maryland
College Park, MD 20742

Leslie Kitchen
Computer Science Center
University of Maryland
College Park, MD 20742

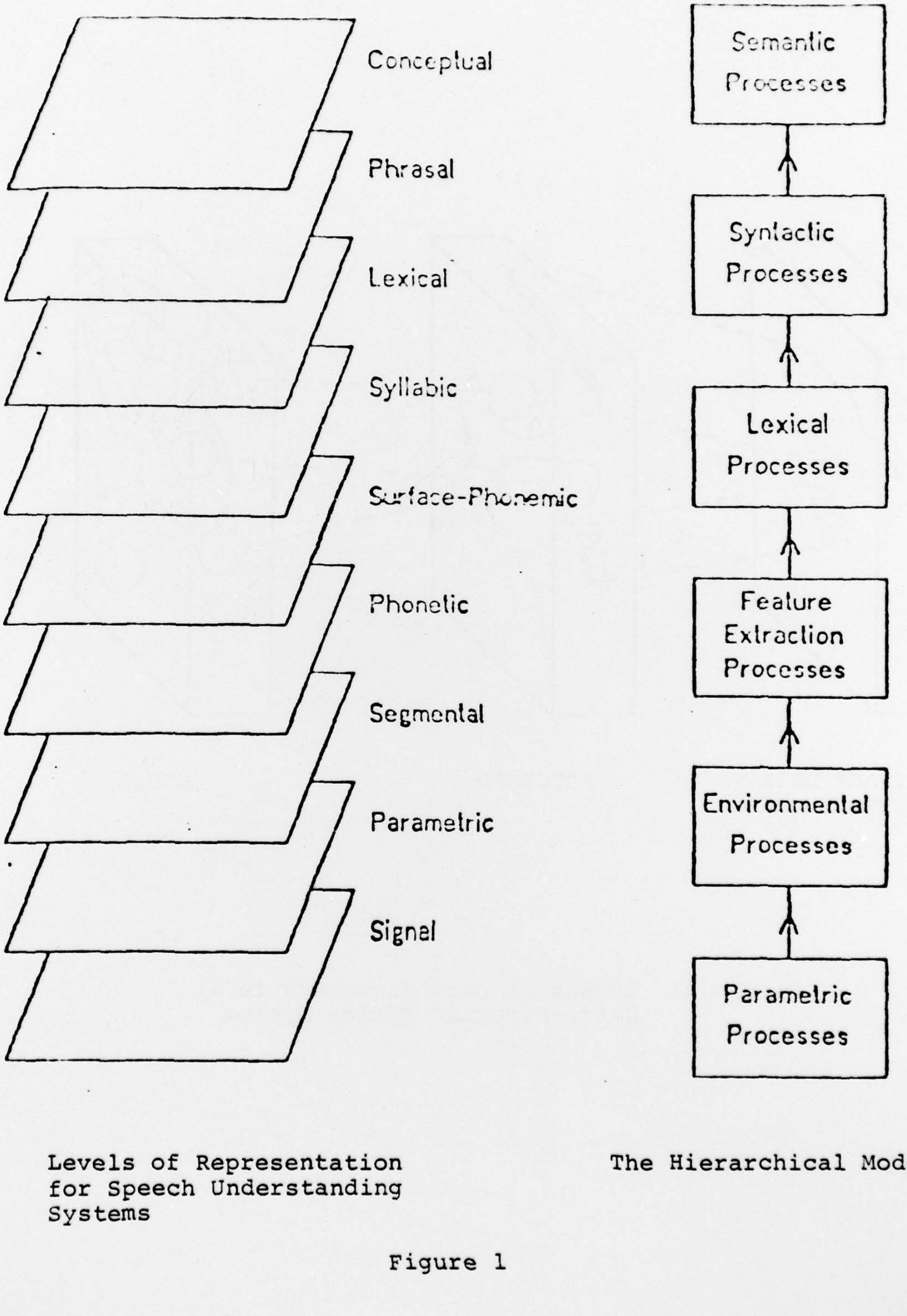
Cesare C. Parma
Department of Computer & Information Science
University of Massachusetts
Amherst, MA 01003

Shmuel Peleg
Computer Science Center
University of Maryland
College Park, MD 20742

Steven Shafer
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213

Steven Small
Department of Computer Science
University of Maryland
College Park, MD 20742

David R. Smith
Department of Computer Science
Carnegie-Mellon University
Pittsburgh, PA 15213



Levels of Representation for Speech Understanding Systems

The Hierarchical Model

Figure 1

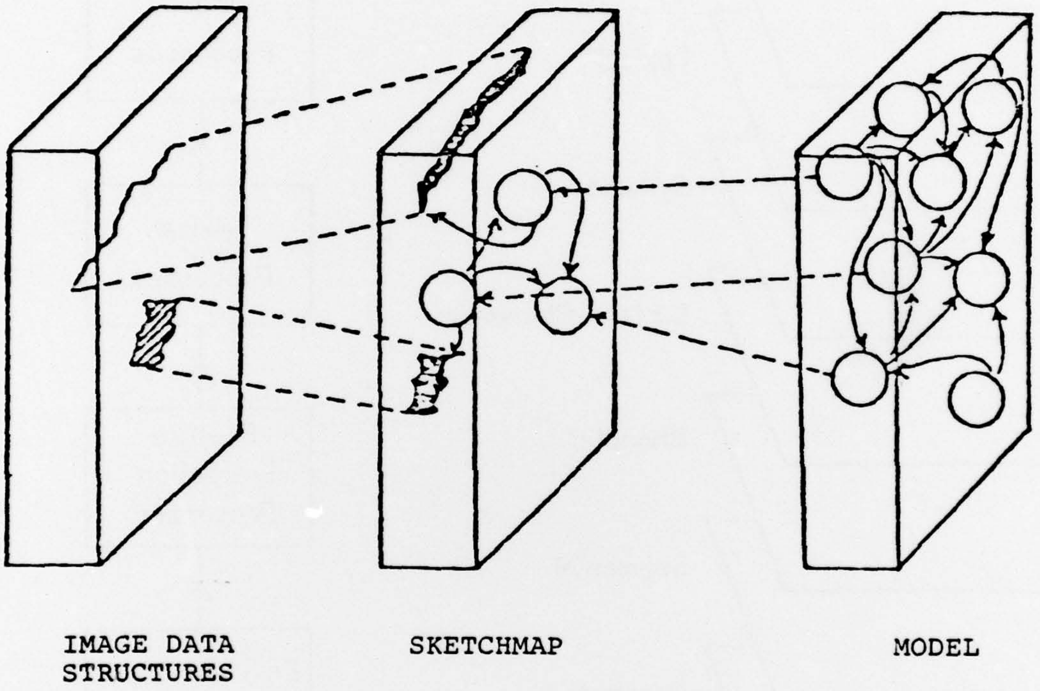


Figure 2. Levels of Data Structure in a
Query-Directed Vision System

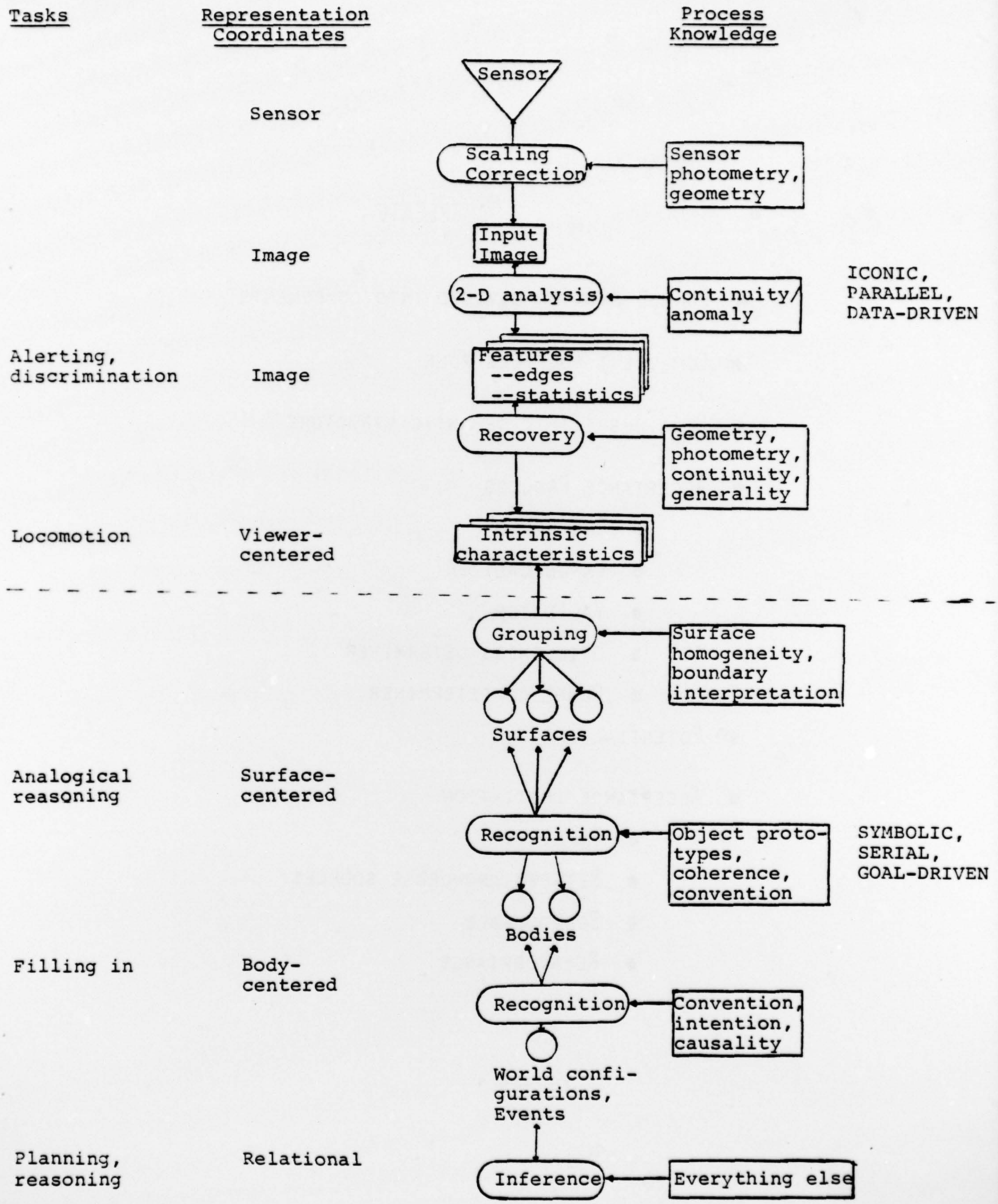

Tasks
Representation Coordinates
Process Knowledge
Sensor
Sensor
Scaling Correction
Sensor photometry, geometry
Input Image
Image
2-D analysis
Continuity/ anomaly
ICONIC, PARALLEL, DATA-DRIVEN
Alerting, discrimination
Image
Features --edges --statistics
Recovery
Geometry, photometry, continuity, generality
Locomotion
Viewer-centered
Intrinsic characteristics
Grouping
Surface homogeneity, boundary interpretation
Surfaces
Analogical reasoning
Surface-centered
Recognition
Object proto-types, coherence, convention
SYMBOLIC, SERIAL, GOAL-DRIVEN
Bodies
Filling in
Body-centered
Recognition
Convention, intention, causality
World confi-gurations, Events
Planning, reasoning
Relational
Inference
Everything else

Figure 3

## AGGREGATE

- Formed from or divided into components
- Component fulfills role
- Domain-specific semantic structure
- Acceptance Process
  - Integrator
  - Canonicalizer
  - Validator
  - Uniqueness determiner
  - Conflict determiner
- Potential units
- Acceptance initiation
  - Explicit
  - Between knowledge sources
  - Schedulable
  - Re-acceptance

Figure 4

# HYPOTHESIS

- PLACEHOLDER FOR A DELAYED DECISION
  -- EVENTUALLY REPLACED BY SELECTED ALTERNATIVE

- REASON FOR DELAY IS TO OBTAIN GUIDENCE FOR SELECTION

- ALTERNATIVES
  - EXPLICIT OR IMPLICIT (GENERATOR)
  - MUST BE ON SAME BLACKBOARD LEVEL AS HYPOTHESIS
  - NORMALLY AGGREGATES, BUT MAY BE OTHER HYPOTHESES

- SELECTION
  - EXPLICIT OR IMPLICIT (PREFERENCE ORDER) CHOICE
  - ASSUME OR DEDUCE MODE

Figure 5

## CONTEXT

- CREATED BY 'ASSUME' OPERATION (AS A CHILD CONTEXT)
- KNOWLEDGE SOURCES ARE SATISFIED AND EXECUTED WITHIN A CONTEXT
- PERMITS PARALLEL NON-INTERFERING (NON-SHARED) EXPLORATIONS
- AUTOMATIC INHERITANCE ACROSS DIRECTED, BOOLEAN GRAPH OF CONTEXTS
- EXPLICIT MECHANISMS FOR MOVING INFORMATION FROM CHILD TO PARENT (ANCESTOR) CONTEXT

Figure 6

# SCHEDULING

- SCHEDULING KNOWLEDGE-SOURCES REACT TO CHANGES ON SCHEDULING BLACKBOARD (SUCH AS NEW ACTIVATION RECORDS)
- SCHEDULING KNOWLEDGE SOURCES MUST THEMSELVES BE SCHEDULED
- USER-PROVIDED BASE SCHEDULER BREAKS RECURSION
- ONLY BASE SCHEDULER CAN EXECUTE AN ACTIVATION RECORD (VIA INVOKE SUBROUTINE)
- SCHEDULING KNOWLEDGE SOURCES CAN SUGGEST ACTIVATIONS TO EXECUTE
- BASE SCHEDULER MANAGES ACTIVATION SUGGESTIONS

Figure 7

## KNOWLEDGE SOURCE ACTIVATION

STEPS

1. FIRING PATTERN SATISFIED
2. AFTER RUNNING KNOWLEDGE SOURCE COMPLETES
   A. ACTIVATION RECORD CREATED
   B. ACTIVATION PLACED ON SCHEDULING BLACKBOARD AT COMPUTED LEVEL
3. ACCEPTANCE OF ACTIVATION RECORD
4. ACTIVITY TO PLACE IT IN SCHEDULING STRUCTURE(S)
5. SELECTION FOR EXECUTION
6. REVALIDATION
   - FIRING PATTERN
   - FIRING CONTEXT
7. EXECUTION

Figure 8

## PRODUCTION SYSTEMS

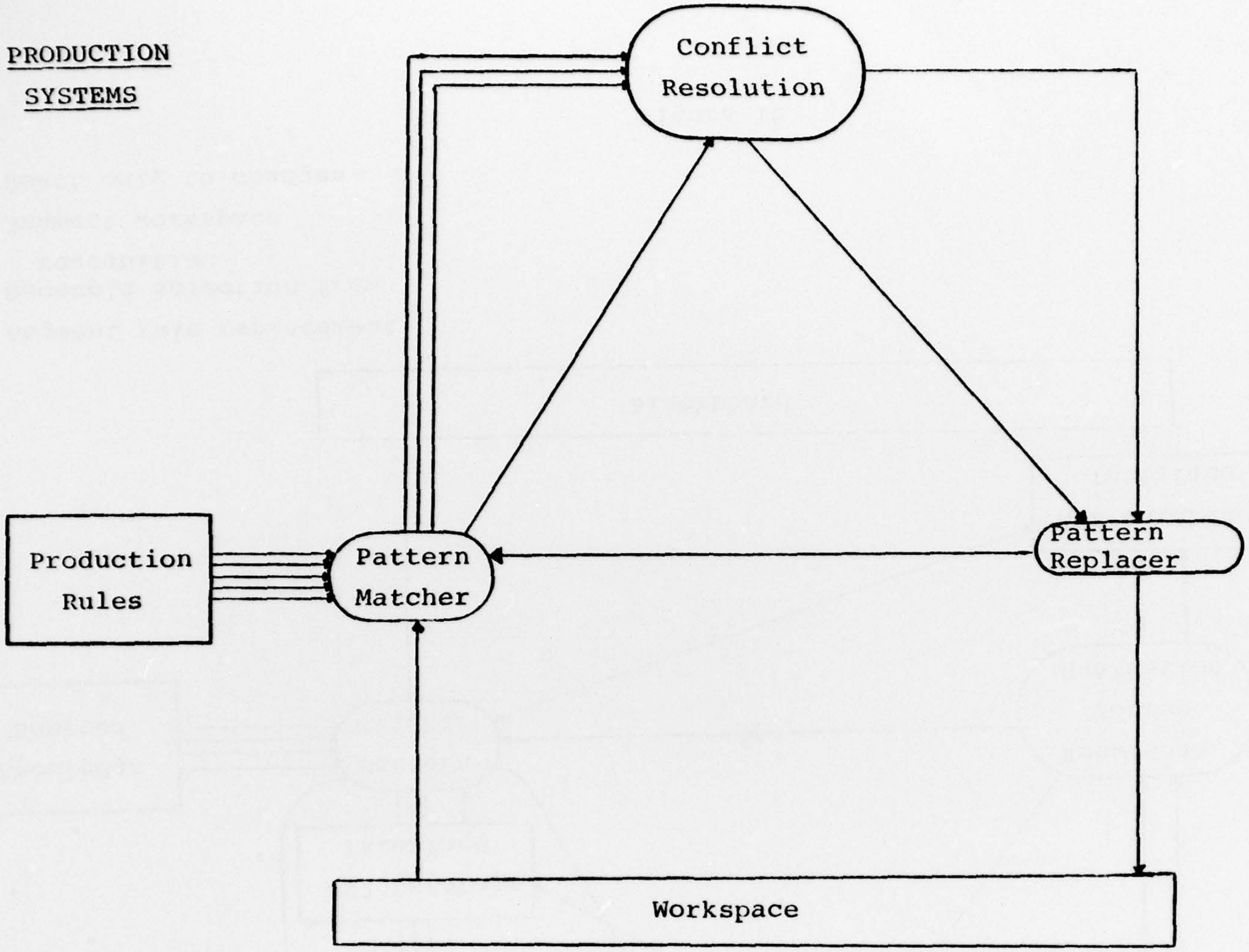
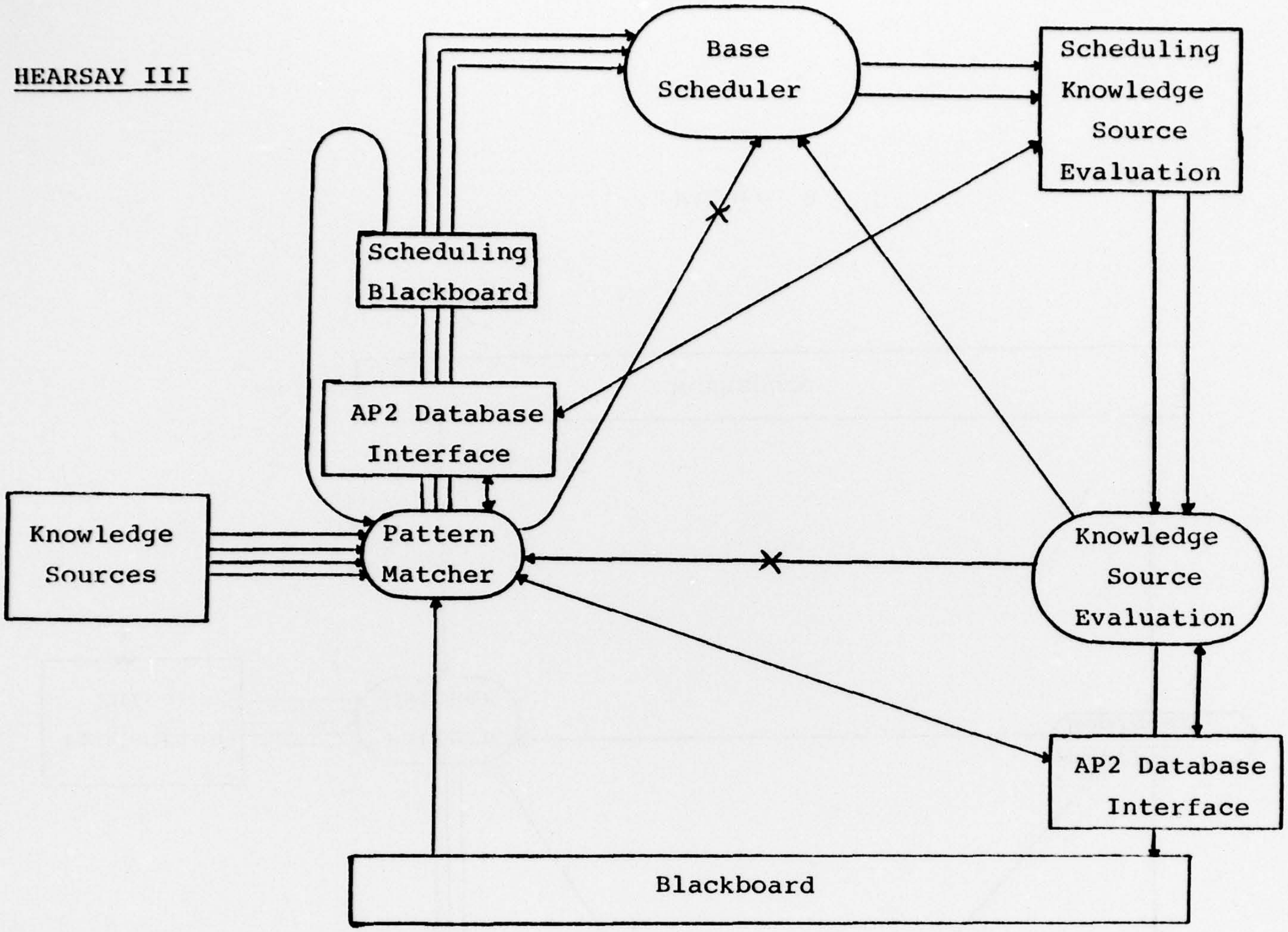


Figure 9

## HEARSAY III



- Augment rule replacement
- Decouple selection from recognition
- Augment workspace
- React only to changes

Figure 10

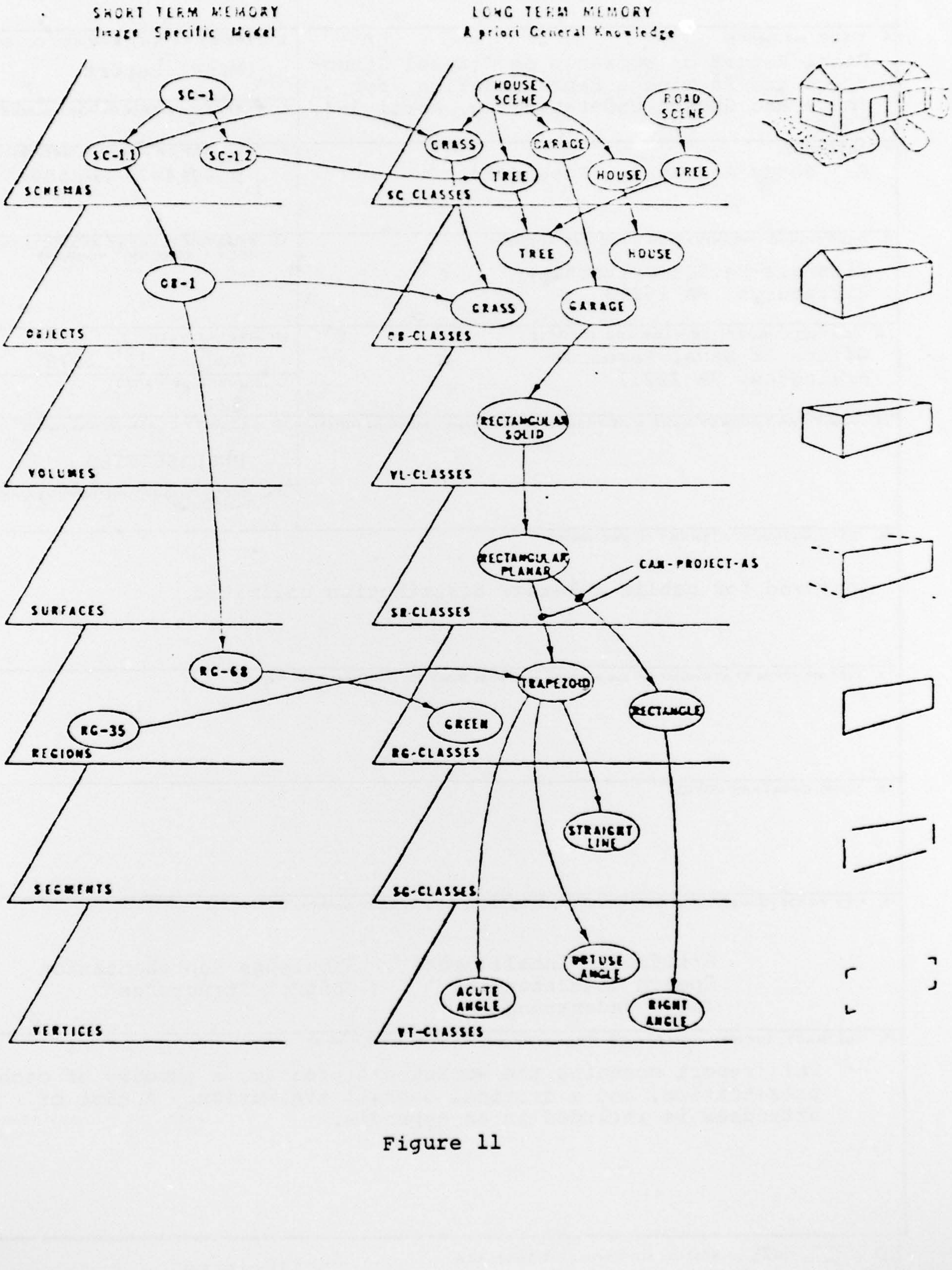

SHORT TERM MEMORY
Image Specific Model
LONG TERM MEMORY
A priori General Knowledge
SC-1
SC-11
SC-12
SCHEMAS
HOUSE SCENE
ROAD SCENE
GRASS
GARAGE
TREE
HOUSE
TREE
SC CLASSES
TREE
HOUSE
OB-1
GRASS
GARAGE
OBJECTS
OB-CLASSES
RECTANGULAR SOLID
VOLUMES
VL-CLASSES
RECTANGULAR PLANAR
CAN-PROJECT-AS
SURFACES
SR-CLASSES
RG-68
TRAPEZOID
RECTANGLE
RG-35
GREEN
REGIONS
RG-CLASSES
STRAIGHT LINE
SEGMENTS
SG-CLASSES
OBTUSE ANGLE
ACUTE ANGLE
RIGHT ANGLE
VERTICES
VT-CLASSES

Figure 11

| | | |
|---|---|---|
| **REPORT DOCUMENTATION PAGE** | | READ INSTRUCTIONS<br>BEFORE COMPLETING FORM |
| 1. REPORT NUMBER | 2. GOVT ACCESSION NO. | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle)<br>Final Report on Workshop on Control Structures and Knowledge Representation for Image and Speech Understanding, April 3-4, 1979 | | 5. TYPE OF REPORT & PERIOD COVERED<br>Final Report<br>6. PERFORMING ORG. REPORT NUMBER |
| 7. AUTHOR(s)<br>Raj Reddy and Azriel Rosenfeld | | 8. CONTRACT OR GRANT NUMBER(s)<br>N00014-79-M-0009 |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS<br>Carnegie-Mellon University<br>Pittsburgh, PA 15213 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS |
| 11. CONTROLLING OFFICE NAME AND ADDRESS<br>Office of Naval Research<br>Arlington, VA 22217 | | 12. REPORT DATE<br>August 31, 1979<br>13. NUMBER OF PAGES<br>52 |
| 14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) | | 15. SECURITY CLASS. (of this report)<br>UNCLASSIFIED<br>15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report)<br>Approved for public release; distribution unlimited. | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) | | |
| 18. SUPPLEMENTARY NOTES | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number)<br>Artificial Intelligence<br>Speech Understanding<br>Image Understanding<br>Knowledge Representation<br>Control Structures | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number)<br>This report contains the workshop's program, a summary of each presentation, and a critical overall evaluation. A list of attendees is included in an appendix. | | |